# Local Directional Ternary Pattern for Facial Expression Recognition

Byungyong Ryu, *Student Member, IEEE,* Adín Ramírez Rivera, *Member, IEEE,* Jaemyun Kim, *Student Member, IEEE,* and Oksam Chae, *Member, IEEE*

***Abstract*—This paper presents a new face descriptor, local directional ternary pattern (LDTP), for facial expression recognition. LDTP efficiently encodes information of emotion-related features (*i.e.*, eyes, eyebrows, upper nose, and mouth) by using the directional information and ternary pattern in order to take advantage of the robustness of edge patterns in the edge region while overcoming weaknesses of edge-based methods in smooth regions. Our proposal, unlike existing histogram-based face description methods that divide the face into several regions and sample the codes uniformly, uses a two level grid to construct the face descriptor while sampling expression-related information at different scales. We use a coarse grid for stable codes (highly related to non-expression), and a finer one for active codes (highly related to expression). This multi-level approach enables us to do a finer grain description of facial motions, while still characterizing the coarse features of the expression. Moreover, we learn the active LDTP codes from the emotion-related facial regions. We tested our method by using person-dependent and independent cross-validation schemes to evaluate the performance. We show that our approaches improve the overall accuracy of facial expression recognition on six datasets.**

***Index Terms*—face descriptor, local pattern, expression recognition, edge pattern, face recognition.**

## I. Introduction

THE automatic recognition of emotion has been an important field in computer vision since its application in marketing, for example, in the registration of the purchasers responses to certain products [1], [2]. One of the key techniques for recognizing emotions automatically is facial expression recognition, which detects and analyzes human emotions from facial images.

Facial expressions can be represented by appearance changes on the face. Consequently, describing them exactly is the key issue in facial expression recognition for detecting emotions. There are two main approaches to describe facial images: geometric-feature-based and appearance-feature-based methods [3]. The first represents the facial image by encoding location relations of main facial components, like eyes, nose, mouth, etc. [4]–[11]. It can describe the facial image efficiently using a few features, and is invariant to scale and rotation. However, the recognition performance relies on the exact locations of key facial components, which are difficult to detect under appearance changes on the face according to facial expressions [12]. The appearance-feature-based methods can avoid this problem innately. They represent the facial image by using image filters which are applied on the whole-face (holistic) or specific-face regions (local) to extract the appearance variations of facial images. In this category, there are many holistic methods, such as Eigenfaces [13], Fisherfaces [14], 2D PCA [15], LDA [16], and IDA [17]. Though those holistic methods have been used successfully for face recognition, local approaches like LBP [18], [19], LPQ [20], LDP [21], LDN [22], [23], HoG [24] and LTP [25] have been studied broadly in facial expression recognition because local ones, unlike holistic methods, are capable of accommodating local variations that occur by expression changes. In particular, methods that extract edge-based local features and histogram representation [21], [22], [24] proved successful in facial expression recognition as emotion-related facial features have prominent gradient magnitudes. Thus, their histogram representation is robust to small location and code errors.

However, these edge-based local methods with histogram representation still have problems. Extracting edge-based local features in the smooth regions of the face image makes unstable patterns which are sensitive to noise and contribute negatively to the classification result. Spatial information of the face features plays an important role in the expression recognition, but histogram representation is inefficient to preserve spatial information. To increase spatial information in the histogram representation, the number of uniform regions should be increased. However, increasing it will accumulate the sampling error for the codes with lack of samples which will result in decreasing overall recognition performance [26]. The performance degradation is more significant at the smooth region where no prominent and stable edge patterns exist.

In this paper, we propose a new face descriptor, Local Directional Ternary Pattern (LDTP), for facial expression recognition. Motivated by the high edge responses in the boundaries of the emotion-related facial features, we extract edge directional patterns in a face image, while avoiding to generate ones from smooth regions (meaningless for expression recognition), by using the magnitude of the edge response. Specifically, we extract two main edge directions as directional patterns at each local pixel and utilize them to extract a local feature only if the edge response is higher than a threshold determined from experiments. To encode the validation and sign information of an edge direction, we add a ternary pattern to each directional pattern. Moreover, we propose a way to select active edge patterns which have significant accumulation for histogram and positional variation among facial expressions. Based on

Byungyong Ryu, Jaemyun Kim, and Oksam Chae are with the Department of Computer Engineering, Kyung Hee University (emails: read100mn@khu.ac.kr, jaemyunkim@khu.ac.kr, and oschae@khu.ac.kr).

Adín Ramírez Rivera is with the Institute of Computing, University of Campinas (email: adin@ic.unicamp.br).

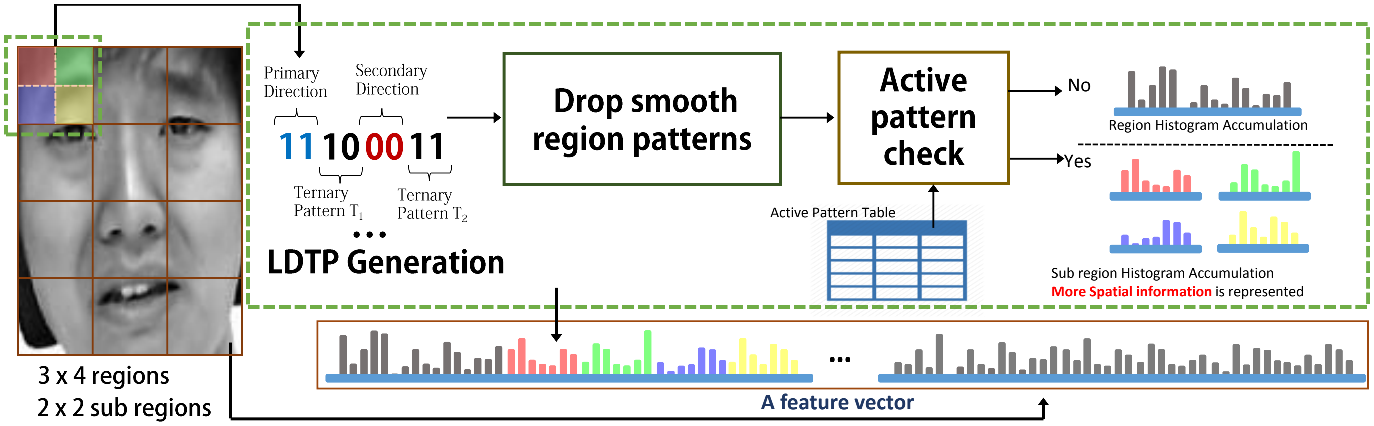


Fig. 1. The overall process of the proposed method. For each face image, we first calculate LDTP codes by using edge response and drop codes from smooth areas. Then, for each region, we add positional bits to active patterns and generated codes are accumulated into a histogram. Lastly, all histograms are concatenated into a feature vector for facial expression recognition.

the selected active edge patterns, we propose a new coding scheme that increases spatial information while suppresses the sampling error, which results in better classification performance in overall recognition. The proposed coding scheme assigns positional bits only to the active edge patterns with significant accumulation to get the effect of using finer grid for the selected codes. It increases the overall performance by applying the finer grid to the active patterns (that require more spatial information and of which sampling error is less significant due to its high accumulation). Figure 1 illustrates the overall process of our proposed method. Experimental results show that our approach improve the performance of facial expression recognition especially in person-independent environments ($N$-person cross validation).

The contributions of our proposal are summarized as follows:

1) We propose a new face descriptor for facial expression recognition. Our method encodes edge directional information of emotion-related features efficiently by removing the meaningless ones from smooth regions in the computed directional patterns.
2) We learn the active edge patterns from the emotion-related regions with substantial accumulation in histogram so that they are exploited in our proposed descriptor efficiently.
3) We propose a new coding scheme that increases spatial information with the active edge pattern, and consequently improves the classification result.
4) We tested our method on six widely used databases to prove improvement of performance of facial expression recognition. The experimental results show that our approach outperforms existing methods.

This paper is structured as follows: In Section II we introduce our method coding scheme. Then, in Section III we describe how to add spatial information to the previously introduced descriptor for facial expression recognition. We carry out experiments of the proposed method and discuss its results in Section IV. Lastly, we conclude our method in Section V.

$$\underset{M_0}{\begin{bmatrix}-1&0&1\\-2&0&2\\-1&0&1\end{bmatrix}}\quad\underset{M_1}{\begin{bmatrix}0&1&2\\-1&0&1\\-2&-1&0\end{bmatrix}}\quad\underset{M_2}{\begin{bmatrix}1&2&1\\0&0&0\\-1&-2&-1\end{bmatrix}}\quad\underset{M_3}{\begin{bmatrix}2&1&0\\1&0&-1\\0&-1&-2\end{bmatrix}}$$
$$\underset{M_4}{\begin{bmatrix}1&0&-1\\2&0&-2\\1&0&-1\end{bmatrix}}\quad\underset{M_5}{\begin{bmatrix}0&-1&-2\\1&0&-1\\2&1&0\end{bmatrix}}\quad\underset{M_6}{\begin{bmatrix}-1&-2&-1\\0&0&0\\1&2&1\end{bmatrix}}\quad\underset{M_7}{\begin{bmatrix}-2&-1&0\\-1&0&1\\0&1&2\end{bmatrix}}$$

Fig. 2. Robinson compass masks used for computing the directions on LDTP.

## II. Local Directional Ternary Pattern

Our proposed Local Directional Ternary Pattern (LDTP) is an eight bit pattern code assigned to each pixel of an input face image. In expression recognition, the shape of the facial features that change according to expressions is more influential than whole-face textures used in face recognition, and the boundaries of the facial features have high edge magnitudes. Therefore, we adopt Robinson compass masks [27] (shown in Fig. 2) as an edge operator to calculate edge responses efficiently, and take two main directions at each pixel to represent local edge shapes. Our method, unlike existing methods, distinguishes useful directional patterns and meaningless ones by using edge magnitude to avoid generating useless patterns, and avoids noise in the smooth regions.

### A. Coding Scheme

The properties of Local Directional Ternary Pattern (LDTP) are: 1) the gradient direction, instead of the gradient magnitude or intensity, is used for superior representation of shapes of the emotion-related facial features; 2) the Robinson compass mask [27] is efficient due to its symmetry; 3) the ternary pattern encodes edge-sign information, and differentiates between edge and smooth (non-edge) regions (thus, solving the weakness of edge patterns in smooth areas).

In the proposed coding scheme, we generate LDTP code by using the edge responses calculated with Robinson compass masks, $\{M_0, \ldots, M_3\}$, we encode the primary and secondary directions, and their corresponding ternary patterns. Robinson compass masks are symmetric and generate the same magnitude response with different signs in opposite directions. For example, in Fig. 2, $M_0$ and $M_4$ masks have same edge response values, with the exception of sign. Therefore, we

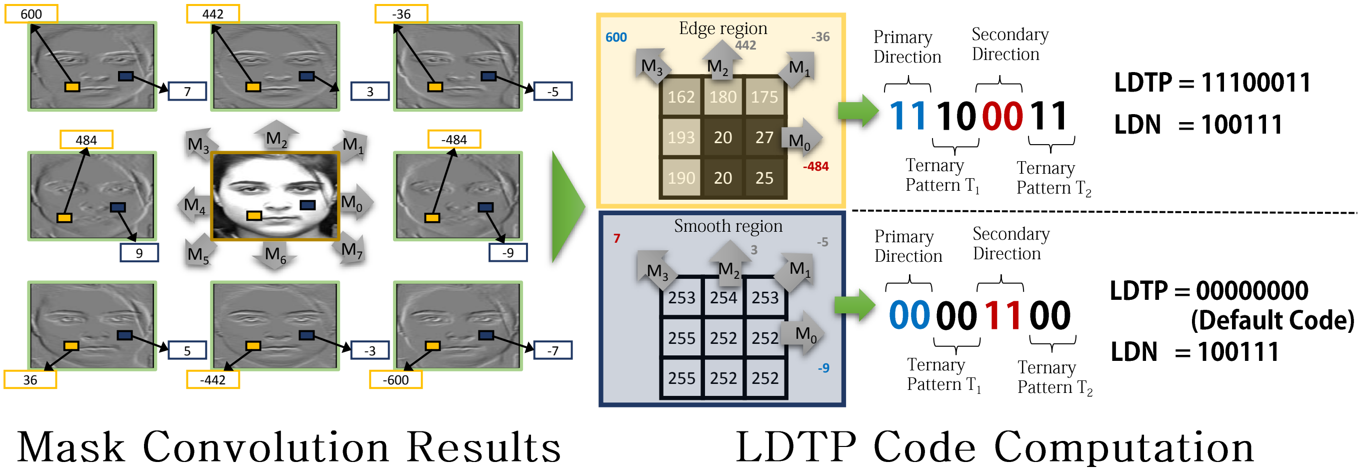


Fig. 3. LDTP code computation. We calculate the edge response by Robinson compass masks with the original image, and select the primary and secondary direction to encode the shape of facial features. We show an example of two 3 × 3 image patches corresponding to the edge responses. LDTP can detect smooth regions by making a different code while other gradient-based patterns cannot (like LDN), as they produce the same code for remarkably different texture.

can use only four masks from $M_0$ to $M_3$ to find the principal directions, which can reduce calculation time. Moreover, as Lahdenoja *et al.* [28] showed, the patterns with high symmetry level occur more frequently in face images. Due to its symmetry, Robinson compass mask can effectively represent the symmetric facial features. We encode that representation by using four directional codes and the sign information to form a ternary pattern.

Unlike the original ternary pattern used in LTP [25], our ternary pattern represents additional information of the principal directions by signaling three conditions (*i.e.*, positive or negative strong edge response and weak edge). Since we use a symmetric mask and the ternary patterns contain the sign information (plus response magnitude) to distinguish edge and smooth regions, we only need half of the compass mask, $\{M_0, \ldots, M_3\}$ for encoding directions. We assign 2 bits to encode the primary directional number, and 2 bits for the secondary one; and each directional number has 2 bits for each ternary pattern, as shown in Fig. 3. The Robinson compass mask is applied over the entire image producing a set of response magnitudes correlated with the four directions:

$$R_i = M_i * I, \qquad 0 \leq i \leq 3, \tag{1}$$

where $I$ is the original image, $M_i$ is the $i$th Robinson compass mask, and $R_i$ is the $i$th response image. Then, we search for the $j$th maximum absolute value $D_j$ of the four Robinson compass mask's responses, defined by:

$$D_j(x,y) = \arg\max_i^j \{|R_i(x,y)| : 0 \leq i \leq 3\}, \tag{2}$$

where $\arg\max_i^j$ is an operator that returns the index $i$ of the $j$th maximum value in the set. As stated before, we will search for the first and second directions, *i.e.*, $j \in \{1,2\}$. We also convert the corresponding principal direction response into a ternary pattern. This operation encodes edge response using three levels (negative, equal, and positive). Additionally, it involves the sign of the edge response. In other words, the ternary pattern indicates whether the direction is located in an edge or in a smooth area. Formally, we encoded it as:

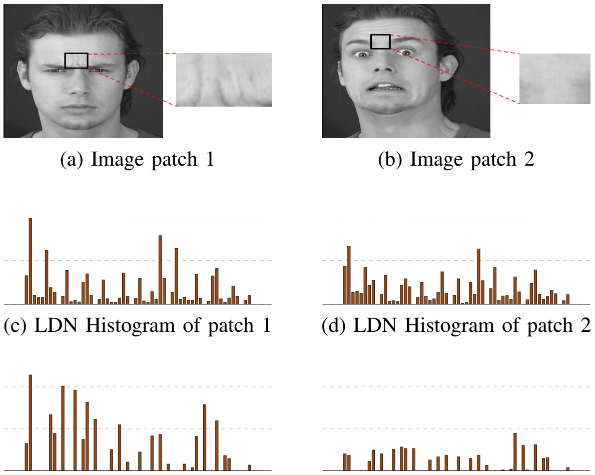
$$T_j(x,y) = \begin{cases} 2 & \text{if } R_{\hat{i}}(x,y) < -\sigma, \\ 1 & \text{if } R_{\hat{i}}(x,y) > \sigma, \\ 0 & \text{if } -\sigma \leq R_{\hat{i}}(x,y) \leq \sigma, \end{cases} \tag{3}$$

where $T_j$ is the ternary pattern of the magnitude of the $j$th direction at position $(x,y)$, $R_{\hat{i}}(x,y)$ is the edge response of $\hat{i}$th direction at position $(x,y)$, $\hat{i} = D_j(x,y)$ is the $j$th principal direction at position $(x,y)$, and $\sigma$ is a threshold value (in our experiments we selected $\sigma$ adaptively as explain in Section IV-A1). The threshold divides the data only in three sections, *upper*, *lower*, and *in between*. However, we can interpret the values as follows, *upper* and *lower* means a strong positive or negative edge response, respectively, whereas *in between* means a weak edge response. With this differentiation, it is possible to separate and keep the directional information from the edges response and ignore the directional information of the smooth areas.

Based on $T_j$, we will determine if $D_j$ is going to be used or not. Hence, for each direction $D_j(x,y)$ the rule is different. If the ternary pattern from the first direction is 0, that means that the pixel $(x,y)$ exists on a smooth area and therefore an empty code (0) is generated which later is being ignored. That is, we check the empty code from the LDTP codes generated in the face image, and do not accumulate them into the histogram—when we are in the face description stage as shown Fig 1. If the ternary pattern from the second direction is 0, only the information of the first direction is meaningful and the knowledge of the second direction can be discarded.

Consequently, the code is created by concatenating the binary form of the two principal directions and the two ternary patterns. This concatenation can be represented by the following operation:

$$\begin{aligned} \text{LDTP}(x,y) = {} & 2^6 D_1(x,y) + 2^4 T_1(x,y) + \\ & 2^2 D_2(x,y) + T_2(x,y), \end{aligned} \tag{4}$$



(a) Image patch 1 (b) Image patch 2

(c) LDN Histogram of patch 1 (d) LDN Histogram of patch 2

(e) LDTP Histogram of patch 1 (f) LDTP Histogram of patch 2

Fig. 4. Two face regions (a) with and (b) without movement during facial expressions, with their respective histograms, (c) and (d), produced by LDN [22] for each patch. And (e) and (f) the histograms of the LDTP method for each patch, respectively. Note that the histograms of LDTP are different while the ones from LDN are similar, despite both regions presenting different characteristics.

where LDTP$(x, y)$ is the code for each pixel $(x, y)$ in the face image, $D_1$ and $D_2$ are the direction number of the primary and secondary directions (from 0 to 3) from the two highest mask responses of the neighborhood of the pixel $(x, y)$, and $T_1$ and $T_2$ are the first and second ternary patterns of the two directions, respectively. An example of the code generation is shown in Fig. 3.

### B. Discriminability of LDTP

Since our method utilizes edge magnitudes and encodes ternary patterns to discard useless ones, it is more discriminative than existing edge-based methods which only encode directions. For example, Figure 4 shows two small image patches of faces with different expressions. There are many edges in the first image patch, Fig. 4(a), due to the expression movement. On the contrary, few edges occur in the second image patch, Fig 4(b), as they are not critical for the expression that is performed. We calculated LDN [22] histogram as an existing edge-based method, and LDTP histogram as shown in Fig. 4. In LDN, the two histograms, Figs. 4(c) and 4(d), are very similar (have no discrimination) since it generates arbitrary patterns on the smooth areas. However, in case of LDTP, the two histograms, Figs. 4(e) and 4(f), are completely distinct due to its filtering process on smooth areas.

## III. Face Description for LDTP

### A. Problems of Histogram-based Face Description

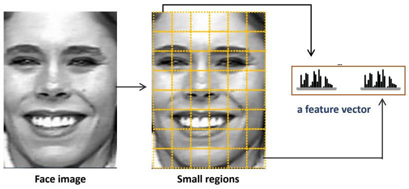
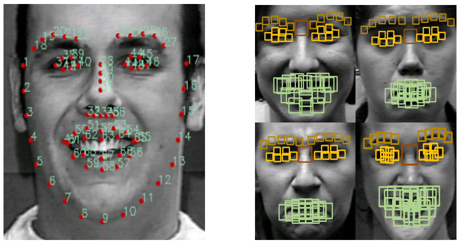


Fig. 5. Creation of histogram based on regions of a face.



(a) (b)

Fig. 6. (a) CK+ AAM 68 landmarks and (b) landmark blocks of emotion-related features.

Many appearance based methods [19], [21], [22], [24], [25], [29]–[31] have used statistical face descriptions as feature vector by using histograms. In this description, the face image is divided into small regions, $\{R^1, ..., R^N\}$, and a histogram $H^k$ of each region $R^k$, which has as many bins as their own pattern codes or image intensities. Thus, the histogram is created based on such region by

$$H^k(c) = \sum_{(x,y)\in R^k} \delta\left(P\left(x,y\right),c\right), \quad \forall\, c, \tag{5}$$

$$\delta(a,b) = \begin{cases} 1 & a = b, \\ 0 & a \neq b, \end{cases} \tag{6}$$

where $c$ is a pattern code, $(x, y)$ is a pixel position inside the region $R^k$, and $P(x, y)$ is a computed pattern code in $(x, y)$. Finally, a global face description is calculated by concatenating those histograms from those regions, as shown in Fig. 5.

This histogram-based description is simple and robust to location and code errors in a small region. However, it needs sufficient sample codes [26], and loses spatial information inside each region—although each region itself contains positional information. Therefore, if we divide the face into more regions, spatial information increases but samples codes in each region decrease due to the smaller size of the regions. Since those insufficient sample codes may lead to unstable histograms (which are called sampling errors), the histogram-based description has a limitation to increase spatial information.

### B. Active Patterns

In the common histogram-based description, the spatial information is extracted from a 2D regular grid placed in the face. However, this strategy is inefficient as it assigns equal importance to all facial features spatial information. Spatial information of emotion-related features is far more critical than others. Instead, we find LDTP codes appearing frequently on emotion-related facial features, and assign more spatial information to them. We call these LDTP codes active.

In this paper, we used tracking data of facial features provided with CK+ [32] to select regions around emotion-related facial features. CK+ presents 68 landmarks tracked by Active Appearance Model (AAM) [33], [34] in every frame [shown in Fig. 6(a)]. Based on EMFACS [35], [36], we utilized

42 emotion-related landmarks on eyes, eyebrows, upper nose, and mouth for selecting active patterns. And we set 17 × 17 pixel block size around each landmark, except on upper nose [the 28th landmark in Fig. 6(a)] around which 33 × 33 pixels block size is set [shown in Fig. 6(b)]. Finally, we calculated LDTP codes (4) in each block and accumulated them into four histograms according to each facial feature (eyes, eyebrows, upper nose, and mouth), as shown in Fig. 7. First, we extract most frequently occurring patterns from the four histograms by

$$d_n^r = \arg\max_c^n \{|\mathbb{H}^r(c)| : c \in \mathbb{LDTP}\}, \tag{7}$$

where $d_n^r$ is the set of the $n$th maximum $c$ patterns (codes) in the $r$th region (*i.e.*, $r \in \{\text{eyebrows}, \text{eyes}, \text{upper nose}, \text{mouth}\}$), $\arg\max_c^n$ is an operator that returns the $n$th maximum value $c$ patterns in the histogram $\mathbb{H}^r$ for the $r$th region, and $\mathbb{LDTP}$ is the set of valid LDTP patterns. We merge $d_n^r$ into a single set by

$$\mathbb{D}_n = \bigcup_{r=0}^{R} d_n^r, \tag{8}$$

where $\mathbb{D}_n$ is the set of unique active patterns comprised by $n$th top patterns from each region. Statistically, the prominent codes will be contained in $\mathbb{D}_n$ without duplication. For example, if we assume that two regions $r = \{0, 1\}$, $n = 2$, $d_2^0 = \{1, 2\}$, and $d_2^1 = \{2, 3\}$ are the resultant sets, $\mathbb{D}_2$ ended up with $\{1, 2, 3\}$ where 2 is on both sets. In this case, we cannot expect a fixed number of active patterns based on $R$ and $n$ alone due to duplications. For securing the expected codes stably, we extract other codes by

$$\mathbb{S}_n = \arg\max_c^j \{|\mathbb{H}^m(c)| : c \in \mathbb{LDTP}, c \notin \mathbb{D}_n\}, \tag{9}$$

where $\mathbb{S}_n$ is a set of the $j$th maximum patterns from the mean histogram $\mathbb{H}^m$, defined by

$$\mathbb{H}^m(i) = \frac{1}{R}\sum_{r=0}^{R-1} \mathbb{H}^r(i), \quad \forall\, i, \tag{10}$$

where $R$ is the number of regions, without containing $\mathbb{D}_n$, $j$ is the number of codes to select, defined by

$$j = Rn - |\mathbb{D}_n|, \tag{11}$$

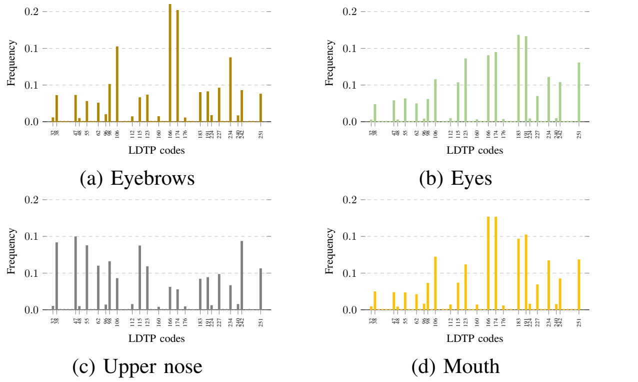
where $R$ is the number of regions ($R = 4$), and $|\mathbb{D}_n|$ is the number of elements of the set. For example, when $n = 4$ and $R = 4$, we expect 16 active patterns but if the number of elements $|\mathbb{D}_n| = 10$, we extract $j = 6$ more codes from mean histogram. And we get the final set of active patterns $\mathcal{D}_n$ by

$$\mathcal{D}_n = \mathbb{D}_n \cup \mathbb{S}_n. \tag{12}$$

According to $n$, we get different active patterns. For example, when $n = 4$, sixteen active patterns are determined, $n = 2$, eight, and $n = 1$, four—assuming $R = 4$.

### C. Face Description for LDTP

In our face description, we generate the code Local Directional Ternary Pattern (LDTP) of each region shown in Fig. 5. Since spatial information of active LDTP codes is more influential to facial expression recognition, we divide a local region into sub-regions of which each has an unique label. And we add more spatial information to active LDTP codes by combining the positional label, as shown in Fig. 8 by



(a) Eyebrows (b) Eyes

(c) Upper nose (d) Mouth

Fig. 7. LDTP histograms on 4 facial features: (a) eyebrows, (b) eyes, (c) upper nose, and (d) mouth.

$$\text{LDTP}_{NM}^n(x, y) = \begin{cases} 2^8 l_{x,y} + c_{x,y}, & c_{x,y} \in \mathcal{D}_n, \\ c_{x,y}, & c_{x,y} \notin \mathcal{D}_n, \end{cases} \tag{13}$$

where $\text{LDTP}_{NM}^n(x, y)$ is the code for each pixel $(x, y)$ in the region divided into $N \times M$ sub regions with certain active LDTP codes $\mathcal{D}_n$, $l_{(x,y)}$ is the label of $N \times M$ sub regions labeled from 0 to $(NM - 1)$, $c_{x,y}$ is the LDTP code at $(x, y)$ calculated by (4), and $\mathcal{D}_n$ is determined by (12). If we do not divide the region into sub regions or $\mathcal{D}_n$ is an empty set, $\text{LDTP}_{NM}^n$ is the same as LDTP. Then, we create the histogram $H^k$ by using (5) and $\text{LDTP}_{NM}^n$.

By employing sub regions and active patterns, we can describe spatial information more efficiently than existing histogram based descriptors. For example, assume that LDTP has 50 different values in which 10 values are active patterns and 40 are normal pattern. Let us divide the face image into 2 × 3 regions with 2 × 2 sub regions. The number of active patterns is 240 (2 × 3 × 2 × 2 × 10) and for normal patterns is 240 (2 × 3 × 40), resulting in 480 dimensions because we do not consider sub regions when we calculate normal patterns as descriptor. However, in existing histogram based description, the dimension is 1200 (2 × 3 × 2 × 2 × 50).

In other words, we can assign more spatial information to active patterns alone by using sub regions, since active patterns are sensitive to position information but normal ones are not. Moreover, active patterns are robust to sampling error since they have a high accumulation by definition. Thus, our approach is reasonable and describe spatial information more efficiently than existing histogram based descriptors.

Finally, global $\text{LDTP}_{NM}^n$ histogram (GLH) is calculated by concatenating all histograms

$$\text{GLH} = \Big\Vert_{k=1}^{K} H^k \tag{14}$$

where $\Vert$ is the histogram concatenation operator, $K$ is the number of regions into which each face image is divided (shown in Fig. 5), and $H^k$ is the histogram computed by (5) using

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2726010, IEEE Transactions on Image Processing

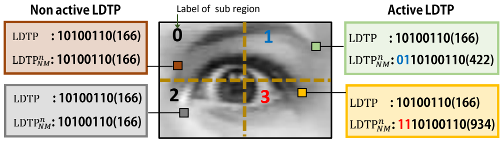

Non active LDTP
Label of sub region
Active LDTP
LDTP : 10100110(166)
$\mathrm{LDTP}^n_{NM}$: 10100110(166)
LDTP : 10100110(166)
$\mathrm{LDTP}^n_{NM}$: 10100110(166)
0
1
2
3
LDTP : 10100110(166)
$\mathrm{LDTP}^n_{NM}$: 0110100110(422)
LDTP : 10100110(166)
$\mathrm{LDTP}^n_{NM}$: 1110100110(934)

Fig. 8. Example of adding more spatial information to active LDTP codes by combining the sub region label.

$\mathrm{LDTP}^n_{NM}$ (13) instead. This GLH is utilized as feature vector representing face image for facial expression recognition.

## IV. Experiments

We executed experiments of facial expression recognition to validate the efficiency of the proposed method. We evaluated our algorithm by using six famous databases: CK+ [32], JAFFE [37], MMI [38], [39], CMU-PIE [40], [41], GEMEP-FERA [42], and BU-3DFE [43]. Basically, we cropped faces from all database images and normalized them to $110 \times 150$ pixels, according to the positions of eyes and mouth provided by the ground truth or the manual selections.

We made use of dependent and independent cross-validation testing schemes, herein called $N$-fold and $N$-person cross-validation, respectively, to evaluate the performance of facial expression recognition accurately. In the $N$-fold cross-validation, we randomly partitioned the image set into $N$ groups. We used $N-1$ groups as the training set and the other group was used as the test set. A large number of methods [21], [44]–[50] adopted this technique with the test data-sets consisting of the three most expressive image frames and easily achieve cross validation. However, as the three most expressive image frames of the same subject are very similar, as shown in Fig. 9, if these frames are partitioned into train and test simultaneously, the classification task will be influenced by the presence of the subject instead of the expression. Some methods [19], [22] add a person-independent restriction in $N$-fold strategy for facial expression recognition. In this strategy, there are various policies on how to support the independence. However, most papers do not reveal the exact details of the person-independent policy in $N$-fold strategy; thus, making the meaning of "person-independent" slightly different. In our experiments, we add $N$-person cross validation [50] in which we exclude one person out of the training set and test it. This strategy always ensures person-independence itself and can more accurately evaluate the performance of facial expression recognition.

The tests were carried out by using SVM [51] with Radial Basis Function (RBF) kernel. As SVM generates binary decisions, multi-class classification can be achieved by using the one-against-one or one-against-all approaches. In this paper, we adopt the one-against-one method, referring to a detailed comparison of multi-class SVMs conducted by Hsu *et al.* [52], which shows that one-against-one method is a competitive approach. For selecting the parameters, a grid-search on the hyper-parameters in a cross-validation approach suggested by Hsu *et al.* [52] was used and the parameter setting with the best cross-validation result was picked.

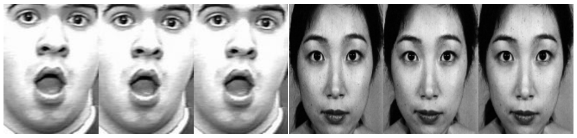


Fig. 9. The three most expressive image frames from CK+ and JAFFE.

### A. Extended Cohn-Kanade Results

The Extended Cohn-Kanade Facial Expression (CK+) database [32] comprises 593 image sequences (from neutral to apex) of 123 subjects who were instructed to perform a series of 23 facial displays. From these sequences, 327 out of 593 have each of the seven basic emotion categories: anger, contempt, disgust, fear, happiness, sadness and surprise. In our setup, we selected 327 sequences with 7 emotion categories. The three most expressive image frames were selected from each sequence to make the 7-class expression dataset (981 images).

*1) Optimal LDTP and $\mathrm{LDTP}^n_{NM}$ Parameters:* The recognition performance of the proposed LDTP and $\mathrm{LDTP}^n_{NM}$ can be affected by its threshold value $\sigma$ in (3), the parameter $n$ for active pattern in (12), sub regions $NM$ in (13), and the number of regions $K$ in (14).

If we use a threshold, $\sigma$, for all face images, the number of included pixels for each face description is different according to image contents. To solve this problem, we select, for each image, the optimal threshold adaptively based on a ratio of included pixels, $\rho$. To select the threshold, we count the amount of pixels among the weak edge responses in the principal direction, $R_{\hat{i}}$, (we store them in a histogram), and choose the minimum value (as the bin) that produces the desire ratio of pixels. Formally, we define the threshold, $\sigma$, based on the ratio of included pixels, $\rho$, as

$$\sigma = \arg\min_{b} \frac{\sum_{i=0}^{b} H_R(i)}{\sum_{j=0}^{1020} H_R(j)} \geq \rho, \tag{15}$$

where $b$ is the bin value in the histogram of edge responses $H_R$ which accumulated histogram has an amount of pixels that is greater or equal to the ratio $\rho$. The histogram is defined (based on all possible pixels) as

$$H_R(c) = \sum_{(x,y)} \delta\big(R_{\hat{i}}(x,y), e\big), \qquad \forall\, e, \tag{16}$$

where $R_{\hat{i}}$ is the edge response in the principal direction $\hat{i} = D_1(x,y)$, $\delta$ is defined as (6), and $0 \leq e \leq 2040$ are the possible edge responses—note that the maximum edge response is $255 \times (1+2+1) \times 2 = 2040$ due to the definition of the Robinson mask. For example, when we set the ratio as $\rho = 0.5$, the threshold $\sigma$ of each image is determined adaptively to include $50\%$ of pixels in the face description. Therefore, we performed an experiment to find the optimum ratio value instead of the optimal threshold $\sigma$. We first tested LDTP with twenty ratios values, $\rho \in \{0.05, 0.10, 0.15, \ldots, 1.0\}$, on $6 \times 7$ regions, as shown in Fig. 10. Considering these recognition rates, we selected the ratio $\rho = 0.7$ as the optimal value to determine the adaptive threshold, $\sigma$, to each image in the following experiments.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2726010, IEEE Transactions on Image Processing

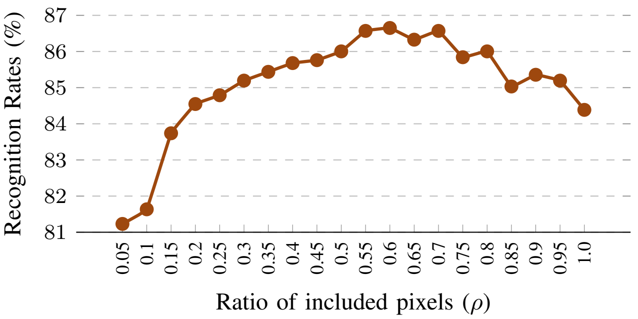


Fig. 10. The recognition rate of LDTP by varying the ratio of included pixels ($\rho$) on 6 × 7 regions in CK+ database.

For the optimal parameters of LDTP$^n_{NM}$, initially, we defined four active patterns with different $n$ values $\{2, 4, 6, 8\}$ in (12) and 3 different sub regions $2 \times 2$, $2 \times 1$, and $1 \times 2$ and tested them to find the optimal active pattern. Next, we searched for the optimal divided regions of LDTP and LDTP$^n_{NM}$ with the found $\sigma$, $n$, and aforementioned three sub regions. All experiments to find the optimal parameters were carried out on CK+ with 7-class and $N$-person cross validation.

We examined the performance of LDTP$^n_{NM}$ with $1 \times 2$ sub regions by varying $n$ $\{2, 4, 6, 8\}$ and the number of regions $\{6 \times 4, 8 \times 5, 10 \times 6\}$ in order to find the optimal active pattern. Figure 11 shows LDTP$^n_{NM}$ recognition results of the different $n$ on several region divisions. As compared with the others, $n = 4$ has the best result in recognition performance. Accordingly, we determine the optimal active pattern as $\mathcal{D}_4$ and set $n$ of active pattern as 4 in succeeding experiments.

We tested the performance of LDTP and LDTP$^n_{NM}$ for the facial expression recognition by varying the number of divided regions with the optimal $\sigma$ and $n$ value in order to find the optimal number of regions. We set three LDTP$^4_{12}$, LDTP$^4_{21}$, LDTP$^4_{22}$ by setting $n = 4$ with the three sub regions. For LDTP, the number of divided regions ranges from $2 \times 4$ to $14 \times 16$, LDTP$^4_{12}$ from $2 \times 2$ to $14 \times 8$, LDTP$^4_{21}$ from $1 \times 4$ to $7 \times 16$, and LDTP$^4_{22}$ from $1 \times 2$ to $7 \times 8$ with consideration of total sub-regions $NMK$ to fit the number of regions tested for LDTP, as shown in Table I. In more detail, because LDTP$^4_{12}$ divides a region into two sub-regions vertically, we determine the vertical number of its regions as a half of LDTP. LDTP$^4_{21}$ and LDTP$^4_{22}$ also have the same context except reducing the horizontal number or both the horizontal and vertical number.

Table I shows the LDTP, LDTP$^4_{12}$, LDTP$^4_{21}$, and LDTP$^4_{22}$ recognition results of the different number of divided regions. A small value of regions results in a lower recognition rate. A larger value of regions increases the performance but after a certain value, decreases. This observation agrees with the problem of histogram-based description in which sufficient sample is needed for effective description. Based on our observation, we selected $8 \times 10$ regions as the optimal number of regions for LDTP. In following experiments, we set $8 \times 10$ regions for LDTP results. LDTP$^4_{12}$ shows the best result in $10 \times 6$ regions while it has better recognition results than LDTP. Since LDTP$^4_{12}$ assigns double vertical spatial information to active patterns $\mathcal{D}_4$ by dividing a region into two sub-regions vertically, $10 \times 6$ regions can be matched to $10 \times 12$ of LDTP which is more than the optimal number ($8 \times 10$) of regions for LDTP. This observation shows that LDTP$^4_{12}$ can encode spatial information more efficiently with enough sample codes accumulated into histograms. Contrasting LDTP$^4_{12}$, LDTP$^4_{21}$ results are not better than those of LDTP. One of the main reasons is that movements of facial parts have principally vertical directions [35], [36]. The recognition results of LDTP$^4_{22}$ have the best result in $5 \times 6$ regions with better results than LDTP. This observation indicates that LDTP$^4_{22}$ also is more efficient than LDTP. Consequently, in following experiments, we use LDTP$^4_{12}$ with $10 \times 6$ regions and LDTP$^4_{22}$ with $5 \times 6$ regions while dropping LDTP$^4_{21}$.

TABLE I
7-CLASS FACIAL EXPRESSION RECOGNITION RATE (%) OF LDTP, LDTP$^4_{12}$, LDTP$^4_{21}$, AND LDTP$^4_{22}$ FOR DIFFERENT NUMBER OF REGIONS ON CK+ DATABASE.

| LDTP | | LDTP$^4_{12}$ | | LDTP$^4_{21}$ | | LDTP$^4_{22}$ | |
|---|---|---|---|---|---|---|---|
| Regions | Results | Regions | Results | Regions | Results | Regions | Results |
| 2 × 4 | 88.07 | 2 × 2 | 88.18 | 1 × 4 | 87.87 | 1 × 2 | 88.28 |
| 4 × 6 | 91.54 | 4 × 3 | 91.44 | 2 × 6 | 90.52 | 2 × 3 | 91.44 |
| 6 × 8 | 92.66 | 6 × 4 | 93.07 | 3 × 8 | 92.35 | 3 × 4 | 93.27 |
| 8 × 10 | **93.58** | 8 × 5 | 93.17 | 4 × 10 | 92.97 | 4 × 5 | 93.68 |
| 10 × 12 | 92.66 | 10 × 6 | **94.09** | 5 × 12 | **93.17** | 5 × 6 | **94.19** |
| 12 × 14 | 91.74 | 12 × 7 | 93.27 | 6 × 14 | 92.0 | 6 × 7 | 93.27 |
| 14 × 16 | 90.62 | 14 × 8 | 93.07 | 7 × 16 | 91.64 | 7 × 8 | 93.17 |

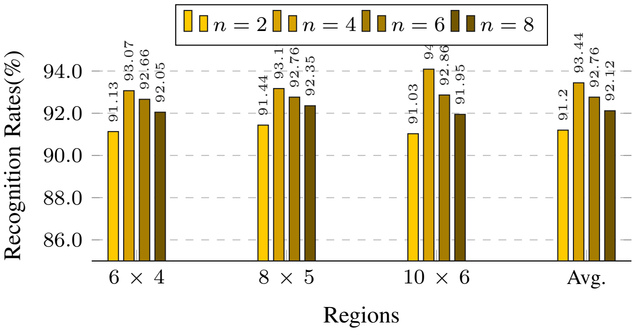
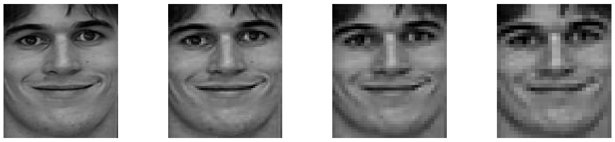


Fig. 11. The recognition rate of LDTP$^n_{NM}$ with 1 × 2 sub regions by varying $n$ and the number of regions in CK+ database.

*2) LDTP vs LDTP$^n_{NM}$:* We compare LDTP$^n_{NM}$ with LDTP for verifying efficiency. For this test, we formed four different resolution face images, $\{110 \times 150, 55 \times 75, 36 \times 48, 27 \times 37\}$, shown in Fig. 12 by down-sampling the $150 \times 110$ images. Our motivation is to evaluate how robust is LDTP$^n_{NM}$ to sample errors, which occur due to lack of sample codes accumulated into the histogram. Then, we tested LDTP, LDTP$^4_{12}$, and LDTP$^4_{22}$ with their optimal parameters as well as LBP implemented based on [19]. All these experiments also were carried out with 7-class prototypical expression dataset of CK+ database and $N$-person cross validation.

As Figure 13 reveals, LDTP$^n_{NM}$ and LDTP has no big difference in $110 \times 150$ and $55 \times 75$ resolutions. However, meaningful differences between LDTP and LDTP$^n_{NM}$ in $27 \times 37$ and $36 \times 48$ resolutions are observed. This indicates that LDTP$^n_{NM}$ is more efficient and robust to the sample errors than LDTP. Also the proposed LDTP, LDTP$^4_{12}$, and LDTP$^4_{22}$ outperform LBP in all resolutions.



(a) 110 × 150 (b) 55 × 75 (c) 36 × 48 (d) 27 × 37

Fig. 12. Four different resolutions of face images from CK+ database.

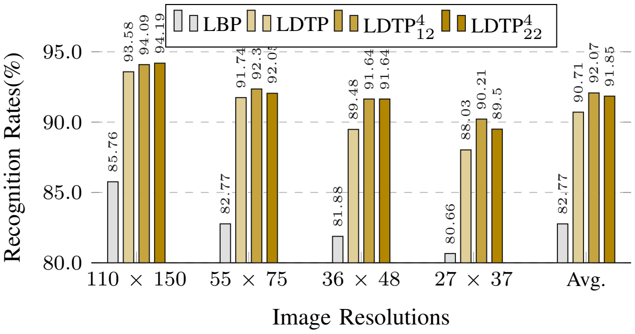


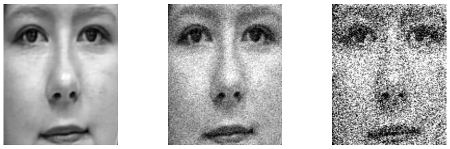
Fig. 13. Comparison of LDTP with $\text{LDTP}^n_{NM}$.

*3) Comparison results:* We tested $\text{LDTP}^4_{12}$ and LDTP by being compared with several existing methods. First, we conducted the facial expression recognition by using the seven expressions dataset and $N$-person cross validation. In this test, we compared $\text{LDTP}^4_{12}$ and LDTP with several geometric and appearance based methods, such as similarity-normalized shape (SPTS) and canonical appearance features (CAPP) reported by Lucey *et al.* [32] when proposing CK+ database, a constrained local model (CLM) based method suggested by Chew *et al.* [53], a CLM method by using shape related information only (CLM-SRI) proposed by Jeni *et al.* [54], a method based on emotion avatar image (EAI) proposed by Yang *et al.* [55], SRC+IVR [56], and DNN [57]. Table II presents that our method performs better than all other methods.

TABLE II
7-CLASS EXPRESSION RECOGNITION ACCURACY (%)FOR DIFFERENT DESCRIPTORS ON CK+ DATABASE (PERSON-INDEPENDENT).

| Descriptor | Accuracy |
|---|---|
| SPTS [32]* | 50.4 |
| CAPP [32]* | 66.7 |
| SPTS+CAPP [32]* | 83.3 |
| CLM [53]* | 74.4 |
| CLM-SRI [54]* | 88.6 |
| EAI [55]* | 82.6 |
| SRC+IVR [56]* | 90.3 |
| DNN [57]* | 93.2 |
| $\text{LDTP}_{\rho=1.0}$ | 89.6 |
| LDTP | 93.6 |
| $\text{LDTP}^4_{12}$ | 94.1 |
| $\text{LDTP}^4_{22}$ | **94.2** |

* Notice that this result is from the corresponding original paper.



(a) No noise (b) 0.08–0.16 (c) 0.16–0.32

Fig. 14. Three different noisy images from CK+ database.

TABLE III
7-CLASS EXPRESSION RECOGNITION ACCURACY (%) USING $N$-PERSON CROSS VALIDATION. GS MEANS 5 × 5 GAUSSIAN SMOOTHING IS APPLIED TO IMAGES BEFORE THE TEST (PERSON-INDEPENDENT).

| Descriptor | No noise | 0.08–0.16 | 0.16–0.32 |
|---|---|---|---|
| LBP [19] | 85.84 | 72.09 | 69.09 |
| LBP [19]+GS | — | 81.23 | 73.62 |
| LDP [21] | 88.07 | 78.28 | 57.44 |
| $\text{LDN}^K$ [22] | 88.58 | 76.45 | 52.75 |
| $\text{LDTP}_{\rho=1.0}$ | 89.56 | 84.71 | 72.90 |
| LDTP | 93.58 | 89.91 | **86.24** |
| $\text{LDTP}^4_{12}$ | 94.09 | **91.13** | 84.30 |
| $\text{LDTP}^4_{22}$ | **94.19** | 90.52 | 85.73 |

Second, we carried out the expression recognition on noisy images by using $N$-person cross validation. We produced two different noisy images by applying Gaussian noise on each image with zero mean and in two random intervals of standard deviations (0.08–0.16, 0.16–0.32). Examples of noisy images can be seen in Fig. 14. In this case, we compared our methods against LBP [19], LDP [21], and LDN with Kirsch masks ($\text{LDN}^k$) [22], which were implemented and set the number of regions same as those described in their papers. In addition, to verify the effectiveness of eliminating edge patterns from smooth area in LDTP, we added to the experimental results LDTP with $\rho = 1.0$ defined in (15) which utilizes all pixels of the face image. Tables III presents the recognition performance of several descriptors produced by $N$-person cross validation on 7-class CK+ dataset. Not just LDTP, but also $\text{LDTP}^4_{12}$ and $\text{LDTP}^4_{22}$, outperform all other methods. Specially, our proposed methods shows outstanding recognition rates in noisy images. $\text{LDTP}^4_{12}$ and $\text{LDTP}^4_{22}$, which encode spatial information more efficiently, accomplish better performance of facial expression recognition than LDTP, demonstrating $\text{LDTP}^n_{NM}$ description is robust to noise as well.

## *B. JAFFE Results*

The Japanese Female Facial Expression (JAFFE) database [37] consists of 213 images of 10 Japanese females. Each image is labeled one of the seven emotions (anger, disgust, fear, happy, sadness, surprise, and neutral) and has a 256 × 256 resolution. When the subject made facial expressions, her hair was tied back for exposing her face area related to expressions. Like CK+ database, we made 7-class expression dataset to contain images labeled neutral, and 6-class expression dataset without neutral.

We tested the proposed method on JAFFE database by using $N$-fold cross validation. In this test, we made comparison of our method with Gabor [58], LDP [21], and SRC+IVR [56]. Table IV shows the comparison results of LDTP and $\text{LDTP}^n_{NM}$. We observed that the recognition rate in JAFFE database has no big difference of that in CK+ shown in Table II. However, JAFFE contains fewer images than CK+ so we cannot ensure sufficient sample image for training, and some expressions are labeled incorrectly or expressed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2726010, IEEE Transactions on Image Processing

TABLE IV
FACIAL EXPRESSION RECOGNITION ACCURACY (%) FOR DIFFERENT DESCRIPTORS ON JAFFE DATABASE USING $N$-FOLD CROSS VALIDATION (PERSON-DEPENDENT).

| Descriptor | 6-class | 7-class |
|---|---|---|
| Gabor [58]* | 85.8 | 80.8 |
| LDP [21]* | 90.1 | 85.4 |
| SRC+IVR [56]* | 94.7 | — |
| LDTP | 93.9 | 93.0 |
| $\text{LDTP}^4_{12}$ | 94.5 | 92.8 |
| $\text{LDTP}^4_{22}$ | **94.8** | **93.2** |

* Notice that this result is from the corresponding original paper.

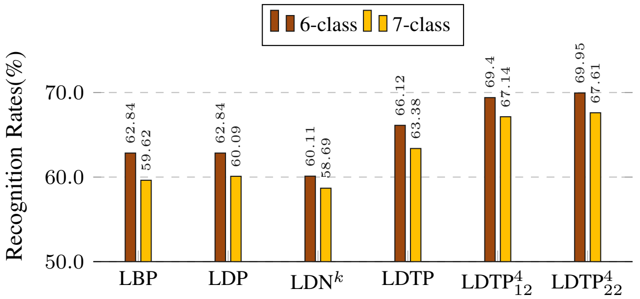


Fig. 15. Facial expression recognition accuracy (%) for different descriptors on JAFFE database using $N$-person cross validation (person-independent).

wrongly [47], [59]. Hence, the results from JAFFE should be far lower than CK+. The main reason of the high recognition rates is that it is likely that face images from the same subject with same expression spread out both training set and test set when we use $N$-fold cross validation with three most expressive image frames.

Therefore, we tested JAFFE database for different descriptors by using $N$-person cross validation for correctly estimating the performance. Furthermore, we tested the JAFFE performance of noisy images which made by the same way as CK+. Figures 15, 16, and 17 show the $N$-person cross validation performance of the recognition of several methods on the 6- and 7-class JAFFE dataset. In $N$-person cross validation of JAFFE, the performance of the recognition is far lower than CK+. This low result is reasonable to us because of aforementioned problems of JAFFE for facial expression recognition. Our methods outperform all other methods both in the noisy images and the normal images. From this observation, we found that $\text{LDTP}^4_{12}$ and $\text{LDTP}^4_{22}$ perform better in the noisy images than LDTP. That is, encoding scheme used in $\text{LDTP}^n_{NM}$ is robust to not only low resolutions but also nosy images.

Additionally, the LDTP, $\text{LDTP}^4_{12}$, and $\text{LDTP}^4_{22}$ confusion matrices for 7-class expression recognition with $N$-person cross validation on JAFFE database are presented in Fig. 18. Disgust, fear, sad emotions are more confused to another. And we found that because neutral recognition rate is similar to the average recognition rate, there is not much difference between 6-class and 7-class.

## C. MMI Results

The MMI face database [38], [39] contains more than 1,500 samples of both image sequences and static images of faces in frontal and profile view with various facial expressions of emotion, including single and multiple facial muscle activation. For our experiments, we used the frontal face images (Part II) which consists of 238 sequences of 28 subjects (sessions 1767 to 2004). Each of the sequences is labeled one of 6 emotions (anger, disgust, fear, happiness, sadness, and surprise). Persons wearing glasses were recorded twice with their glasses and without.

We first tested our methods on MMI database by using $N$-fold cross validation. In this test, we compared our methods with LBP [19], Common Patches (CPL) [60], Common and Specific Patches (CSPL) [60], ADL [60], AFL [60], LDA [61], and DCNN [62]. Table V presents that LDTP, $\text{LDTP}^4_{12}$, and $\text{LDTP}^4_{22}$ perform better than all others. Like JAFFE, we can see the high recognition rates due to $N$-fold cross validation. As images with glasses in MMI database play a bad role in the performance and it has less sample images than CK+, the high recognition rates are wrong. Therefore, we tested our methods

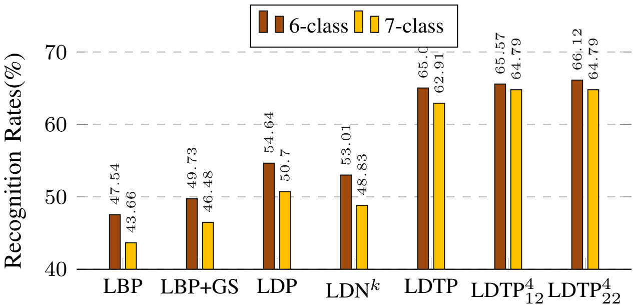


Fig. 16. Facial expression recognition accuracy (%) for different descriptors on JAFFE database with noise (0.08–0.16 standard deviation) using $N$-person cross validation. GS means $5 \times 5$ Gaussian smoothing is applied to images before the test (person-independent).

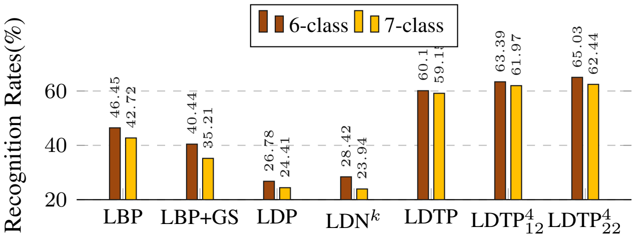


Fig. 17. Facial expression recognition accuracy (%) for different descriptors on JAFFE database with noise (0.16–0.32 standard deviation) using $N$-person cross validation. GS means $5 \times 5$ Gaussian smoothing is applied to images before the test (person-independent).

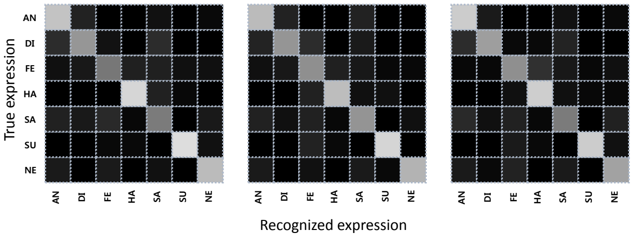


Fig. 18. Confusion matrix of 7-class expression recognition for LDTP (left), $\text{LDTP}^4_{12}$ (center), and $\text{LDTP}^4_{22}$ (right) on JAFFE database using $N$-person cross validation. Note that brighter intensity presents higher scores. The full name of expression labels is AN: anger, DI: disgust, FE: fear, HA: happy, NE: neutral, SA: sad, and SU: surprise.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2726010, IEEE Transactions on Image Processing

TABLE V
6-CLASS EXPRESSION RECOGNITION ACCURACY (%) FOR DIFFERENT DESCRIPTORS ON MMI DATABASE USING $N$-FOLD CROSS VALIDATION (PERSON-DEPENDENT).

| Descriptor | Accuracy |
|---|---|
| CPL [60]* | 49.4 |
| CSPL [60]* | 73.5 |
| AFL [60]* | 47.7 |
| ADL [60]* | 47.8 |
| LDA [61]* | 93.3 |
| DCNN [62]* | 98.6 |
| LDTP | 97.3 |
| LDTP$^4_{12}$ | 99.7 |
| LDTP$^4_{22}$ | **99.8** |

* Notice that this result is from the corresponding original paper.

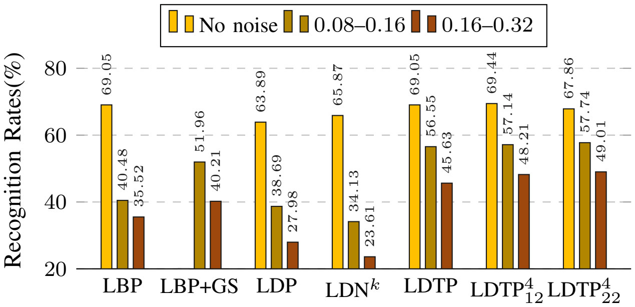


Fig. 19. Facial expression recognition accuracy (%) for different descriptors on MMI database with noise (0.08–0.16 and 0.16–0.32 standard deviation) using $N$-person cross validation. GS means 5 × 5 Gaussian smoothing is applied to images before the test. (person-independent).

on MMI by using $N$-person cross validation for accurate performance of facial expression recognition. Furthermore, we tested the MMI performance of noisy images like CK+. Figure 19 presents the $N$-person cross validation performance of several methods on 6-class MMI dataset. It reasonably shows lower recognition rates (the highest is 69.4%) than CK+ (the highest is 94.19%). LDTP$^4_{12}$ shows the better result on the images without noise but not much different. However, all our methods show big difference of recognition rate on the images with noise. That is to say, effectiveness of eliminating edge patterns from smooth area in LDTP is proved by this observation.

Moreover, we presents the recognition rate of each subject respectively from our LDTP$^4_{12}$ descriptor in Fig. 20 for accurate analysis of the performance. In this observation, we found very low recognition rate on some people who are old and expressed without distinction among expressions as shown in Fig. 21.

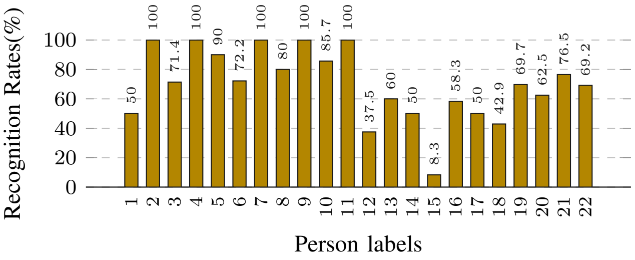


Fig. 20. Facial expression recognition accuracy (%) of each person for LDTP$^4_{12}$ on MMI database using $N$-person cross validation.

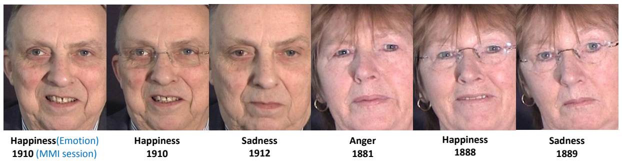


Fig. 21. People expressed without distinction among expressions in MMI database.

### D. CMU-PIE Results

CMU Pose, Illumination, and Expression (CMU-PIE) database [40], [41] contains 41,368 face images of 68 subjects taken under 13 different camera poses, 43 diverse illumination conditions, and with four different expressions (neutral, smile, blinking, and talk). As temporal information is needed to recognize blinking and talking and we deal with the expression recognition for a static image, we tested two expression (neutral and smile) recognition with 27 (center), 05 (right to 27), 29 (left to 27), 07 (below 27) and 09 (above 27) camera poses.

To estimate the performance properly, we tested CMU-PIE database for different descriptors by using $N$-person cross validation. Table VI shows the recognition comparison results on CMU-PIE database. In this observation, we found that the variations of camera angles drop the recognition rate even if it is two-class classification. Note that methods encoding only edge directions such as LDP [21], LDN$^K$ [22], and LDTP are worse than LBP [19]. However, LDTP$^4_{12}$ and LDTP$^4_{22}$ demonstrate the robust performance by outperforming LBP.

### E. GEMEP-FERA Results

The GEMEP-FERA emotion detection dataset [42] contains 134 videos displaying various expressions by 10 subjects, while uttering the sustained vowel ‘aaa’ or meaningless words. In this dataset, 155 videos of 7 subjects are partitioned into the training set, and 134 videos of 6 subjects into the test data, half of which are not included in the training set. Each video is labeled one of the five emotions (anger, fear, joy, relief, and sadness) and has a 720 × 576 resolution. In GEMEP-FERA, the baseline method was provided, which computes LBP features of all frames and then classifies each frame by SVM with a radial basis function kernel. To recognize the emotion label of a video, it calculates and find the emotion label with the maximum occurrence [42].

We tested our method on GEMEP-FERA emotion detection database by using the provided training and test set. In this

TABLE VI
2-CLASS EXPRESSION RECOGNITION ACCURACY (%) FOR DIFFERENT DESCRIPTORS ON CMU-PIE DATABASE USING $N$-PERSON CROSS VALIDATION. (PERSON-INDEPENDENT).

| Descriptor | Accuracy |
|---|---|
| LBP [19] | 88.9 |
| LDP [21] | 88.4 |
| LDN$^K$ [22] | 87.1 |
| LDTP | 89.5 |
| LDTP$^4_{12}$ | 89.3 |
| LDTP$^4_{22}$ | **89.5** |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2726010, IEEE Transactions on Image Processing

TABLE VII
FACIAL EXPRESSION RECOGNITION ACCURACY (%) FOR DIFFERENT DESCRIPTORS ON GEMEP-FERA DATABASE.

| Descriptor | Person independent | Person specific | Average |
|---|---|---|---|
| Baseline [42]* | 44.0 | 73.0 | 56.0 |
| CLM [53]* | 62.0 | 55.0 | 60.0 |
| PHOG+SVM [63]* | 66.7 | 69.0 | 67.0 |
| PHOG+LPQ+SVM [63]* | 64.8 | 83.8 | 72.4 |
| PHOG+LPQ+LMNN [63]* | 62.9 | 88.7 | 73.4 |
| EAI+LPQ [20]* | **75.2** | 96.2 | **83.8** |
| LDTP$_{\rho=1.0}$ | 60.0 | 94.4 | 73.88 |
| LDTP | 63.8 | **98.1** | 77.6 |
| LDTP$^4_{12}$ | 65.0 | 96.3 | 77.6 |
| LDTP$^4_{22}$ | 71.3 | 96.3 | 81.3 |

* Notice that this result is from the corresponding original paper.

test, we also use maximum occurrence mechanism to decide the emotion label of a video. We compared our method against the baseline method [42], CLM [53], PHOG [63], PHOG+LPQ [63], and EAI+LPQ [20]. Further, we added LDTP$_{\rho=1.0}$ to verify the effectness of eliminating edge patterns from smooth regions. Table VII shows comparison results on GEMEP-FERA. Our method outperform other methods except EAI+LPQ which is the FERA (Facial Expression Recognition and Analysis) Challenge [42] ranked 1 method. Although, the proposed methods showed a lower recognition rate than EAI+LQP, our methods are simpler in contrast to avatar image generation and complicated normalization process used in EAI+LPQ. We observed that LDTP$^4_{12}$ and LDTP$^4_{22}$ have a high recognition rate in person-independent tests.

### F. BU-3DFE Results

The BU-3DFE dataset [43] provides 2400 face images with 6 prototype emotions by 100 subjects (100 subjects × 6 expressions × 4 intensities). This dataset has been considered challenging by researchers as its images vary on ethnic/racial ancestries and intensity of expression. Each image is labeled one of the six emotions (anger, disgust, fear, happiness, sadness, and surprise) with an emotion intensity from 01 to 04 and has a 512 × 512 resolution, as shown in Fig. 22.

In our experiments, we selected 2400 images with four level of intensities. We evaluated our method on BU-3DFE database by using $N$-fold (person-dependent) and $N$-person (person-independent) cross validations. Our method was compared with ELM [64], CLM+Median [65], SRC+IVR [56], LBP [19], LDP [21] and LDN$^K$ [22] methods. Additionally, we added LDTP$_{\rho=1.0}$ for the effectness of threshold in LDTP. Table VIII shows the proposed method outperforms other ones. This is because LDTP$^4_{12}$ and LDTP$^4_{22}$ better describe the fine position difference according to expression intensities than other methods. In particular, the difference between LDTP and LDTP$_{\rho=1.0}$ (which utilizes all pixels of the face image) is large in this experiment. This is because the images of BU-3DFE dataset appear as black backgrounds other than the face area as shown in Fig. 22, and LDTP$_{\rho=1.0}$ uses meaningless patterns generated on the black background. Since LDP, LDN, and LBP also generate feature vectors using all the pixels of the face image, they show relatively low results in BU-3DFE dataset.

TABLE VIII
FACIAL EXPRESSION RECOGNITION ACCURACY (%) FOR DIFFERENT DESCRIPTORS ON BU-3DFE DATABASE.

| Descriptor | Person dependent | Descriptor | Person independent |
|---|---|---|---|
| ELM [64]* | 62.9 | LBP [19] | 56.2 |
| CLM+Median [65]* | 76.7 | LDP [21] | 61.3 |
| SRC+IVR [56]* | 87.8 | LDN$^K$ [22] | 56.5 |
| LDTP$_{\rho=1.0}$ | 63.2 | LDTP$_{\rho=1.0}$ | 57.7 |
| LDTP | 86.6 | LDTP | 71.3 |
| LDTP$^4_{12}$ | 87.5 | LDTP$^4_{12}$ | **72.7** |
| LDTP$^4_{22}$ | **88.1** | LDTP$^4_{22}$ | 72.5 |

* Notice that this result is from the corresponding original paper.

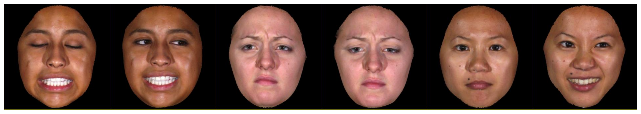


Fig. 22. Facial expression images with black backgrounds in BU-3DFE database.

## V. CONCLUSIONS

In this paper we proposed a new local pattern, LDTP, that efficiently encodes shapes of emotion-related features (*i.e.*, eyes, eyebrows, upper nose, and mouth) by using the directional information. For robust encoding LDTP incorporates ternary patterns that allow it to distinguish directional patterns on edge or smooth regions in which arbitrary, meaningless, and noise-sensitive patterns are generated. For robust facial expression recognition, we describe face image spatially efficiently based on LDTP by using active patterns and sub regions which help our description assign more spatial information to emotion-related facial features. Additionally, we analyzed the performance of facial expression recognition by the use of the two different strategies ($N$-person and $N$-fold cross-validation).

We found that the directional information is suitable to describe shapes of emotion-related facial features, which makes LDTP a more discriminable and robust pattern than existing methods for facial expression recognition. And, we observed that the use of ternary pattern makes the proposed LDTP produce more reliable and stable codes than existing edge-based methods since it removes uncertainty of directional pattern generated in smooth region. Moreover, we studied that our novel face description using active pattern and sub regions gives better performance of facial expression recognition for certain conditions. For instance, the combinations of the active pattern ($n = 4$) and the 1 × 2 or 2 × 2 sub regions (LDTP$^4_{12}$ and LDTP$^4_{22}$) show better ability of facial expression recognition than LDTP with existing histogram based description.

## ACKNOWLEDGMENT

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2015R1A2A2A01006412), and a grant funded by CONICYT, under FONDECYT Iniciación No. 11130098.

## REFERENCES

[1] P. Weinberg and W. Gottwald, "Impulsive consumer buying as a result of emotions," *Journal of Business research*, vol. 10, no. 1, pp. 43–57, 1982.

[2] D. Consoli, "Emotions that influence purchase decisions and their electronic processing," *Annales Universitatis Apulensis: Series Oeconomica*, vol. 11, no. 2, p. 996, 2009.

[3] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in *Handbook of face recognition*. Springer, 2005, pp. 247–275.

[4] H. Hong, H. Neven, and C. von der Malsburg, "Online facial expression recognition based on personalized galleries," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, apr 1998, pp. 354 –359.

[5] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 172–187, jan. 2007.

[6] N. Bourbakis and P. Kakumanu, "Skin-based face detection-extraction and recognition of facial expressions," in *Applied Pattern Recognition*. Springer, 2008, pp. 3–27.

[7] N. Bourbakis, A. Esposito, and D. Kavraki, "Extracting and associating meta-features for understanding peoples emotional behaviour: Face and speech," *Cognitive Computation*, vol. 3, no. 3, pp. 436–448, 2011.

[8] P. Kakumanu and N. Bourbakis, "A local-global graph approach for facial expression recognition," in *Tools with Artificial Intelligence, 2006. ICTAI'06. 18th IEEE International Conference on*. IEEE, 2006, pp. 685–692.

[9] A. Cheddad, D. Mohamad, and A. A. Manaf, "Exploiting voronoi diagram properties in face segmentation and feature extraction," *Pattern recognition*, vol. 41, no. 12, pp. 3842–3859, 2008.

[10] X. Xie and K.-M. Lam, "Facial expression recognition based on shape and texture," *Pattern Recognition*, vol. 42, no. 5, pp. 1003–1011, 2009.

[11] E. Revealed, "Recognizing faces and feelings to improve communication and emotional life," *Henry Holt & Co*, 2003.

[12] M. Pantic and M. S. Bartlett, *Machine analysis of facial expressions*. I-Tech Education and Publishing, 2007.

[13] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[14] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.

[15] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 1, pp. 131–137, 2004.

[16] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *JOSA A*, vol. 14, no. 8, pp. 1724–1733, 1997.

[17] Z. Nenadic, "Information discriminant analysis: Feature extraction with an information-theoretic objective," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 8, pp. 1394–1407, 2007.

[18] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[19] C. F. Shan, S. G. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803 –816, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885608001844

[20] S. Yang and B. Bhanu, "Facial expression recognition using emotion avatar image," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 866–871.

[21] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI journal*, vol. 32, no. 5, pp. 784–794, 2010.

[22] A. Ramirez Rivera, R. Castillo, and O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *Image Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 1740–1752, 2013.

[23] A. Ramirez Rivera and O. Chae, "Spatiotemporal directional number transitional graph for dynamic texture recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[24] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based hog features," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 884–888.

[25] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[26] J. Ylioinas, A. Hadid, Y. Guo, and M. Pietikäinen, "Efficient image appearance description using dense sampling based local binary patterns," in *Computer Vision–ACCV 2012*. Springer, 2013, pp. 375–388.

[27] G. S. Robinson, "Edge detection by compass gradient masks," *Computer Graphics and Image Processing*, vol. 6, no. 5, pp. 492–501, 1977.

[28] O. Lahdenoja, M. Laiho, and A. Paasio, "Reducing the feature vector length in local binary pattern based face recognition," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2. IEEE, 2005, pp. II–914.

[29] T. Gritti, C. Shan, V. Jeanne, and R. Braspenning, "Local features based facial expression recognition with face registration errors," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–8.

[30] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *Image Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 533–544, 2010.

[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[32] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[33] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[34] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[35] W. V. Friesen and P. Ekman, "Emfacs-7: Emotional facial action coding system," *Unpublished manuscript, University of California at San Francisco*, vol. 2, p. 36, 1983.

[36] P. Ekman, E. Rosenberg, and J. Hager, "Facial action coding system affect interpretation dictionary (facsaid)," 1998.

[37] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, apr 1998, pp. 200 –205.

[38] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 5–pp.

[39] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in *Proc. Intl Conf. Language Resources and Evaluation, Wshop on EMOTION*, 2010, pp. 65–70.

[40] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 46–51.

[41] ——, "The cmu pose, illumination, and expression database," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1615–1618, 2003.

[42] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 921–926.

[43] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*. IEEE, 2006, pp. 211–216.

[44] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[45] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Automatic face and gesture recognition, 1998. proceedings. third ieee international conference on*. IEEE, 1998, pp. 454–459.

[46] F. Y. Shih, *Image processing and pattern recognition: fundamentals and techniques*. John Wiley & Sons, 2010.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2726010, IEEE Transactions on Image Processing

[47] X. Feng, M. Pietikainen, and A. Hadid, "Facial expression recognition with local binary patterns and linear programming," *Pattern Recognition And Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii*, vol. 15, no. 2, p. 546, 2005.

[48] X. Fu and W. Wei, "Centralized binary patterns embedded with image euclidean distance for facial expression recognition," in *Natural Computation, 2008. ICNC'08. Fourth International Conference on*, vol. 4. IEEE, 2008, pp. 115–119.

[49] G. Guo and C. R. Dyer, "Simultaneous feature selection and classifier training via linear programming: A case study for face expression recognition," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–346.

[50] F. Cheng, J. Yu, and H. Xiong, "Facial expression recognition in jaffe dataset based on gaussian process classification," *Neural Networks, IEEE Transactions on*, vol. 21, no. 10, pp. 1685–1690, 2010.

[51] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[52] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, mar 2002.

[53] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, and S. Sridharan, "Person-independent facial expression detection using constrained local models," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 915–920.

[54] L. A. Jeni, D. Takacs, and A. Lorincz, "High quality facial expression recognition in video streams using shape related information only," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2168–2174.

[55] S. Yang and B. Bhanu, "Understanding discrete facial expressions in video using an emotion avatar image," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 4, pp. 980–992, 2012.

[56] S. H. Lee, K. N. K. Plataniotis, and Y. M. Ro, "Intra-class variation reduction using training expression images for sparse representation based facial expression recognition," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 340–351, 2014.

[57] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 2016, pp. 1–10.

[58] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction." in *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*, vol. 5, june 2003, p. 53.

[59] F. Ahmed, H. Bari, and E. Hossain, "Person-independent facial expression recognition based on compound local binary pattern (clbp)." *Int. Arab J. Inf. Technol.*, vol. 11, no. 2, pp. 195–203, 2014.

[60] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2562–2569.

[61] J. Wang and L. Yin, "Static topographic modeling for facial expression recognition and analysis," *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 19–34, 2007.

[62] P. Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki, "Dexpression: Deep convolutional neural network for expression recognition," *arXiv preprint arXiv:1509.05371*, 2015.

[63] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using phog and lpq features," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 878–883.

[64] A. Iosifidis, A. Tefas, and I. Pitas, "On the kernel extreme learning machine classifier," *Pattern Recognition Letters*, vol. 54, pp. 11–17, 2015.

[65] Z. Pan, M. Polceanu, and C. Lisetti, "On constrained local model feature normalization for facial expression recognition," in *International Conference on Intelligent Virtual Agents*. Springer, 2016, pp. 369–372.

**Byungyong Ryu** (S'15) received the B.Sc. degree in computer engineering from Kyung Hee University, Yongin-si, Gyeonggi-do, South Korea, in 2010, where he is currently pursuing the Ph.D. degree at the Department of Computer Engineering. His current research interests include facial expression, age, and gender recognition using face images, and image enhancement and medical image processing in dentistry.

**Adín Ramírez Rivera** (S'12, M'14) received his B.Sc. degree in Computer Engineering from San Carlos University (USAC), Guatemala in 2009. He completed his M.Sc. and Ph.D. degrees in Computer Engineering from Kyung Hee University, South Korea in 2013. He is currently an Assistant Professor at the Institute of Computing, University of Campinas, Brazil. His research interests are image enhancement, object detection, expression recognition, and pattern recognition.

**Jaemyun Kim** (S'14) received the B.Sc. degree in computer engineering from Kyung Hee University, Yongin-si, Gyeonggi-do, South Korea, in 2010, where he is currently pursuing the Ph.D. degree at the Department of Computer Engineering. His current research interests include object detection and tracking, pattern recognition, facial analysis, video error detection and concealment.

**Oksam Chae** (M92) received the B.Sc. degree in electronics engineering from Inha University, Incheon, South Korea, in 1977, and the M.S. and Ph.D. degrees in electrical and computer engineering from Oklahoma State University, Stillwater, in 1982 and 1986, respectively. He was a Research Engineer with Texas Instruments Image Processing Laboratory, Dallas, TX, from 1986 to 1988. Since 1988, he has been a Professor with the Department of Computer Engineering, KyungHee University, Gyeonggido, South Korea. His current research interests include multimedia data processing environments, intrusion detection systems, sensor networks, and medical image processing in dentistry. Prof. Chae is a member of the SPIE, Korean Electronic Society (KES), and the Institute of Electronics, Information and Communication Engineers.